Arch Virol (2007) 152: 1665–1678
DOI 10.1007/s00705-007-0992-3
Printed in The Netherlands

Archives of Virology

Sequence determination of variable regions within the genomes of gallid herpesvirus-2 pathotypes

S. J. Spatz[1], **R. F. Silva**[2]

[1] Southeast Poultry Research Laboratory, Agricultural Research Service, United States Department of Agriculture, Athens, Georgia, USA
[2] Avian Disease and Oncology Laboratory, Agricultural Research Service, United States Department of Agriculture, East Lansing, Michigan, USA

Received 18 December 2006; Accepted 23 April 2007; Published online 8 June 2007

Summary

Comparative genomic studies of attenuated and virulent strains of *Gallid herpesvirus 2* (GaHV-2) have identified 6 regions of sequence variability. These regions include the open reading frames (ORFs) encoding UL36 and UL49 and regions devoid of large ORFs (132-bp repeats, *a*-like sequences and the junctions flanking the unique short region). Our data indicate that the carboxyl terminus of UL36 contains regions of heterogeneity that are unique to CVI988-derived attenuated strains. A deletion of the TKSERT domain and a glycine^{245} polymorphism in the UL49 proteins were also identified in these derivatives. Phylogenetic analyses of both UL36 and UL49 sequences indicate that CVI988-derived strains partition differently from other attenuated strains (RM-1 and R2/23), indicating that additional mutations contribute to attenuation. In very virulent and very virulent plus strains a single nucleotide polymorphism (SNP) was identified within the 132-bp tandem repeats. Within the junctions flanking the unique short region, these strains also contain deletions in sequences that are predicted to bind the transcription factor NF kappaB. In some attenuated strains, deletions were also identified in the latency-associated transcript (LAT) promoters and adjacent regions encoding microRNAs. These results indicate that virulence is likely multi-factorial with contributions from both multiple genes and cis-acting sites.

Correspondence: Stephen J. Spatz, Southeast Poultry Research Laboratory, Agricultural Research Service, U.S. Department of Agriculture, 934 College Station Rd, Athens, GA 30605, USA
e-mail: sspatz@seprl.usda.gov

Introduction

Marek's disease (MD) is a highly contagious lymphoproliferative and demyelinating disease of chickens caused by the acute-transforming alphaherpesvirus gallid herpesvirus type 2 (GaHV-2) [4, 31]. Over the past three decades, losses to poultry producers have been greatly reduced through the use of live attenuated vaccines [45, 46]. Originally an antigenically related virus, Meleagrid herpesvirus type 1 (MeHV-1) isolated from turkeys, was used as a vaccine to prevent MD [49]. However, within a decade after its introduction this vaccine lost its efficacy, necessitating the incorporation of more antigenically similar vaccine strains

(GaHV-3) for use in combination with MeHV-1. Although this bivalent vaccine provided some enhanced level of protection, it too lost its efficacy, and by the 1990s the preferred vaccine to protect against very virulent (vv) and very virulent plus (vv+) field strains was an attenuated strain of GaHV-2, the CVI988 or Rispens vaccine strain [27, 35, 44]. Since its isolation and characterization in the early 1970s, this strain has been distributed to vaccine manufacturers worldwide, and numerous derivatives are currently available.

The complete genomic sequence of GaHV-2 (strain Md5) was determined by Tulman et al. in the late 1990s, and currently the sequences of two additional virulent strains (GA and Md11) and three vaccine strains (attenuated GaHV-2 [CVI988], GaHV-3 and MeHV-1) are available [1, 14, 19, 23, 28, 40, 42]. Their protein-coding capacity has also been well defined [24, 30, 37]. Unfortunately, only a handful of genes encoding virulence factors have been identified through the generation of knock-out mutants, predominantly in genes that lack homologues in GaHV-3 and MeHV-1 [31]. In 2004, a comparative whole genomic program was initiated by the Agricultural Research Service of the USDA in order to investigate genetic changes between attenuated and non-attenuated strains of GaHV-2. Bioinformatic analyses of the sequence of the Rispens vaccine strain (CVI988) revealed numerous genetic anomalies relative to virulent strains: SNPs, sequence expansions and deletions [23, 28, 40, 42]. We have previously reported that the majority of these anomalies occur within the long repeat regions and affect the genes encoding RLORF12, RLORF6, 23 kDa and RLORF7 (meq) [41]. However, the comparative genomics study also revealed sequence length polymorphisms within the UL36 and UL49 genes, and in regions devoid of large ORFs: the 132-bp repeats, *a*-like sequences and internal repeat short/unique short (IRS/US) and unique short/terminal repeat short (US/TRS) junctions [40]. The objective of this study was to expand on the prior analyses and determine whether these genes and genomic regions from a collection of GaHV-2 pathotypes (attenuated, mildly virulent, virulent, very virulent and very virulent plus) contribute to the phylogenetic partitioning of the pathotypes.

Material and methods

Viruses and cells

Duck and chick embryo fibroblasts (DEFs and CEFs) were prepared from 11-day-old embryos and maintained at 39 °C under 5% CO_2 in Leibovitz's L-15 medium plus McCoy 5A medium (1:1) supplemented with 2.5% bovine serum and antibiotics. The GaHV-2 strains used in this study were kindly provided by Drs. K. A. Schat (Cornell University), N. Osterrieder (Cornell University), M. Parcells (University of Delaware) and P. Sondermeijer (Intervet). Viruses were mainly propagated in duck embryo fibroblasts. Some virus stocks were propagated on chick embryo fibroblasts using similar conditions as those mentioned above except the use of minimal essential medium supplemented with 8% fetal bovine serum. Table 1 contains a list of these strains, their passage history, host cell type and pathotype classification [47].

Purification of total DNA from GaHV-2-infected cells

DNA was extracted from avian fibroblasts infected with GaHV-2 strains listed in Table 1 using the DNeasy Tissue Kit (Qiagen, Valencia, CA) according to the manufacturer's instruction. Briefly, 200 µl of infected cellular suspension was treated with proteolytic enzymes at 70 °C for 10 min. An equal volume of 100% ethanol was then added to the lysate, and the DNA was purified by mini column chromatography and centrifugation. Columns were washed with ethanol/salt solutions to remove cellular proteins. The DNA

Table 1. The strains of GaHV-2 used in this study, passage history and propagation cell types. Abbreviations for the characterized virulence ranking of the GaHV-2 strains are as follows: *a* attenuated, *m* mild, *v* virulent, *vv* very virulent and *vv+* very virulent plus, *CEF* chick embryo fibroblast and *DEF* duck embryo fibroblast. ζ denotes CVI9888 strain passed 5 times in 15 × 7 chickens

Strains	Passage number	Cell type	Pathotype
CVI988 (Intervet)	27	CEF	a
CVI988-BP5ζ	48	CEF	a
RM-1	40	DEF	a
R2/23 (Md11)	106	CEF	a
CU-2	14	CEF	m
JM/102W	13	DEF	v
RB-1B (ADOL)	23	DEF	v
571	9	DEF	v
549a	8	DEF	vv
595	7	DEF	vv
584a	8	DEF	vv+
648a	8	DEF	vv+
686	8	DEF	vv+

Table 2. The sequences of the primers used in the amplifications of the 6 variable regions along with their coordinates based on the Md5 genome and expected sizes of the PCR products

Region	Primer's name	Sequence	Coordinates	Size
UL36	UL36F	AGGCATTGGCAATGACTGTTGG	79273–79294	
	UL36R	TGAAAGCATGTGGGATCGAGTGG	80608–80630	1357
UL49	UL49F	GACTTACTTGCAGTAGTAGG	111167–111186	
	UL49R	CACGGTACAATAGAAGGTGC	111996–112015	848
a-like	13b	CTTGTAGCTTCCTCCGCCTACG	139969–139990	
	ir79	AGGAAGGTGGTCCGGAAACC	143915–143934	3965
IRS/US	ir46	TGGGAACAGCAATCACTTTCG	152178–152198	
	mdvp08	CTCGTGGCTAACTTATGCAGAG	153873–153894	1716
US/TRS	RB1B13f	CCAGCTCGTCCATTAAACG	163524–163542	
	RB1B14r	ACGGGTTTCCTGACTTGAAC	165913–165932	2408
132 bp repeat	mdvp09F	TCTCCGGTTAGCTCCTGGCATTC	127486–127508	
	mdvp09r	AGAGCCGTGTCAATCGACAACTG	129832–129854	2368

was eluted in 10 mM Tris, pH 8.0, and quantified at 260 nm by spectrophotometry.

Polymerase chain reaction conditions

Six regions within the GaHV-2 genome were targeted for sequence analysis. Table 2 lists the primer sequences, coordinates based on the Md5 sequence (GenBank accession number AF243438) and expected size of the polymerase chain reaction (PCR) products. Amplification reactions included 100 ng of total DNA isolated from infected cells, paired oligonucleotide primers and Platinum Taq polymerase (Invitrogen, Carlsbad, CA). PCR conditions were optimized using an Eppendorf Mastercycler Gradient to determine the optimal annealing temperatures, and 5% DMSO was included in the reaction mixtures to increase product yields. The cycling conditions for the six reactions were as follows: all templates were initially denatured at 94 °C for 2 min and then for 30 sec at 94 °C at the beginning of each cycle (total 40). Primers were allowed to anneal for 30 sec at temperatures that varied per reaction: UL36 (68 °C), UL49 (60 °C), *a*-like (64 °C), IRS/US (56 °C), US/TRS (62.5 °C) and 132 bp (60 °C). The extension temperature was 68 °C for 120, 45, 180, 120, 150, and 150 sec, respectively.

Table 3. The haploid copy number of the 132-bp repeats in GaHV-2 strains that represent the 5 pathotypes. An asterisk indicates a strain containing one copy of the 132-bp repeats with the uncommon thymidine at position 98. A double asterisk denotes both tandem copies contain thymidine at position 98. Derivatives of CVI988 are abbreviated: *INV* intervet, *BP5* back passaged five times in 15×7 birds and *BAC* bacterial artificial chromosome

Strain	Haploid copy #	Virulence
CVI988-INV	2	a
CVI988-BP5	2	a
CVI988-BAC	8	a
RM-1	2	a
R2/23	4	a
CU-2	3	m
JM/102W	2	v
RB-1B-ADOL	2	v
571	2	v
GA	2**	vv
549a	2*	vv
595	2*	vv
Md5	2	vv
584a	2*	vv+
648a	2*	vv+
686	2*	vv+

Amplicon cloning and sequencing

PCR products were cloned into the vector pCR2.1-TOPO using the conditions as defined by the manufacturer (Invitrogen, Carlsbad, CA). All ligation products were transformed into *E. coli* DH5α cells and selected on LB plates containing 100 μg/ml of ampicillin and a topically added cocktail of IPTG (25 μl of 100 mg/ml) and X-gal (20 μl of 50 mg/ml). White colonies were screened using two approaches based on the size of the insert and its stability. Colonies generated from cloning experiments involving PCR product <2.0 kb (UL36, UL49, IRS/US and US/TRS regions) were screened using a conventional plasmid mini-prep procedure. White colonies generated by cloning large PCR products containing the *a*-like sequence or 132-bp repeat region were first screened using colony PCR. After positive identification by visualization on 0.6% 1X TAE agarose gels, colonies were then re-screened using standard miniprep DNA isolation. In both approaches, recombinant vectors containing full-length

inserts were identified via restriction digestion with *Eco*RI and agarose gel electrophoresis.

Cloned PCR products were generally sequenced using oligonucleotides with annealing sites flanking the insert. Recombinant plasmids containing the *a*-like sequences and 132-bp repeat regions were sequenced using a battery of oligonucleotides with staggered annealing sites on both strands. Often, this strategy was supplemented with transposon-mediated sequencing based on Epicentre technology (Epicentre, Madison, WI) in order to determine the sequence of regions containing direct repeats. DNA sequencing was performed at the South Atlantic Area sequencing facility (Athens, GA) and Polymorphic DNA Technologies Inc. (Alameda, CA) using the BigDye terminator cycle sequencing protocol and analyzed on a model ABI-3730 XL DNA Analyzer (Applied Biosystems, Foster City, CA).

DNA sequence analysis

DNA sequences were assembled using the Sequencher program (Gene Codes, Ann Arbor, MI) and manually edited. The final sequence for each region represents that obtained through sequencing of both strands. Sequencing data containing ambiguities and polymorphisms were resolved by resequencing. DNA sequences were maintained and analyzed using Lasergene (DNASTAR, Madison, WI), NCBI Entrez and other web-based tools. Homology searches were conducted using the NCBI programs blastP and PsiBlast with default settings, as well as the web-based program PairwiseFLAG. Multiple alignments of proteins and nucleotide sequences were generated using MUSCLE, MultAlin and MEGA 3.1 [6, 10, 22]. Phylogenetic trees were drawn using Phylip [34]. Repeat elements were determined using the Tandyman program (Los Alamos National Laboratory, NM). Transcription factor binding sites were investigated using the web-based program TFSearch [13]. The sequences of the non-attenuated MDV strains Md5, Md11 and GA used for comparison were obtained from GenBank [23, 28, 42]. Nucleotide sequence data reported in this paper have been submitted to GenBank nucleotide sequence database and have been assigned accession numbers EF526115 through EF526179.

Results

PCR amplification and cloning of six variable regions

A comparative genomic study was initiated to identify genes and cis-acting regulatory sequences that differ among attenuated and virulent strains of GaHV-2. In this study, we identified two genes (UL36 and UL49) within the unique long region and four regions (132-bp repeats, *a*-like sequences, IRS/US and US/TRS junctions) that differ considerably among the pathotypes. In order to investigate the relationship between the genotypes and associated phenotypes, the nucleotide sequences of these genes and regions were determined from the 13 pathotyped strains listed in Table 1. To achieve this, whole cellular DNA isolated from GaHV-2-infected avian fibroblasts was subjected to PCR using a battery of oligonucleotides and varying thermal cycling conditions. Amplification of each of the six regions required optimized magnesium concentrations and annealing temperatures. Most amplification reactions generated a homogeneous product (well-defined band) except those involving the *a*-like sequence. PCR products from some, but not all, of these amplification reactions resolved as broad bands on agarose gels. These results were strain dependent and probably reflect heterogeneity in the *a*-like sequences and not PCR artifacts since each genome contains a diploid number of *a*-like sequences. It is also suspected that the GaHV-2 genomes within infected cells exist as a collection of quasispecies containing varying *a*-like sequences, similar to results reported for human herpesvirus type 6 [7].

All PCR products were first cloned into pCR2.1-TOPO prior to sequencing. Recombinants were screened using standard mini-prep procedures or colony PCR. At least three clones were isolated per cloning event and sequenced. Due to size differences of PCR products generated from a single amplification event involving *a*-like regions, clones containing both high- and low-molecular-weight inserts were processed for sequencing. The sequences of the higher-molecular-weight inserts are represented in Fig. 6.

Intragenic variations: UL36 and UL49

In comparative genomic studies involving CVI988, the predicted translation products of two genes (UL36 and UL49) differed among attenuated and non-attenuated strains. Sequencing of the variable region within the UL36 gene from 13 strains provided additional evidence that the carboxyl terminus is extremely diverse in amino acid composition

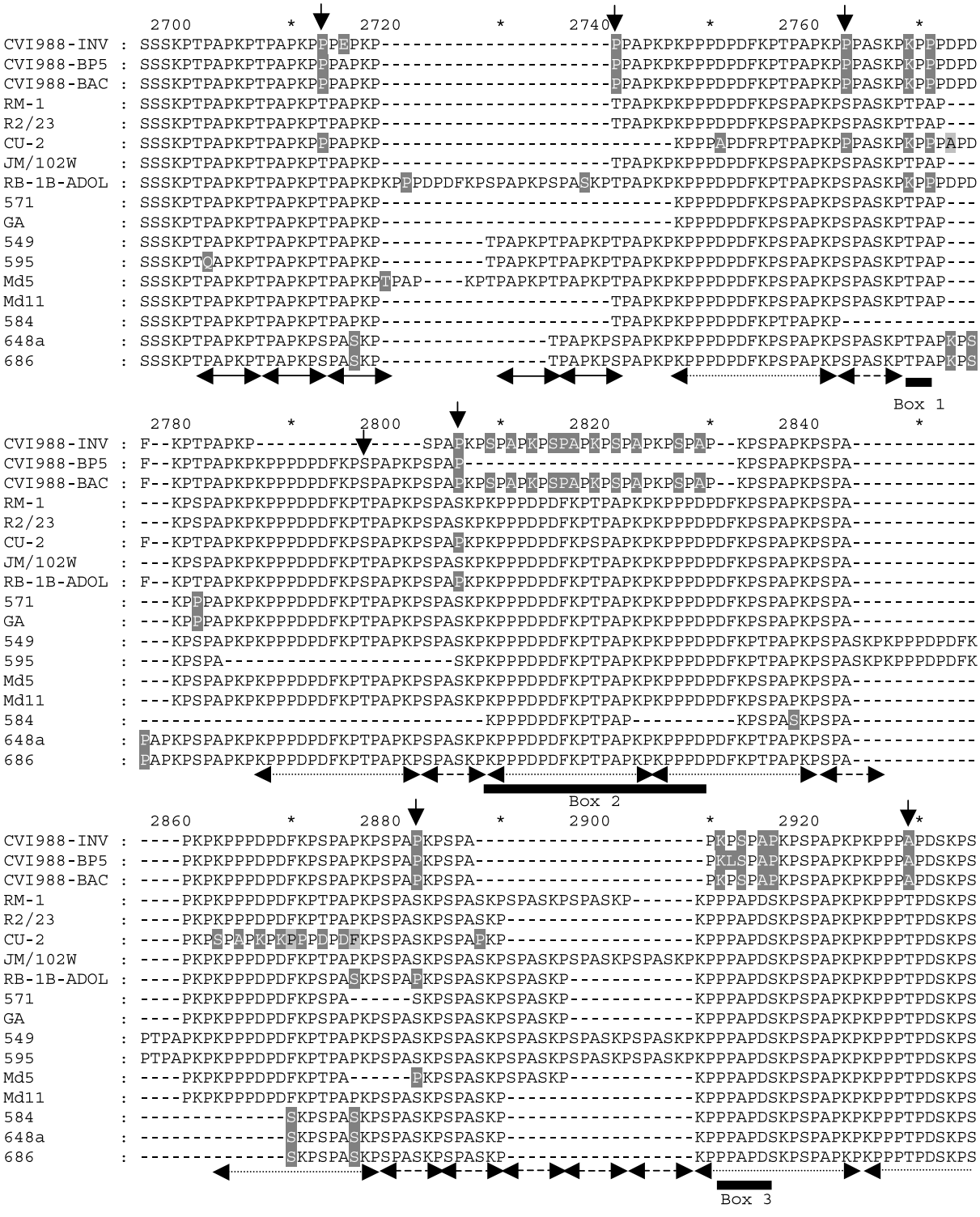

Fig. 1. Multiple sequence alignment of the variable region near the carboxyl terminus of GaHV-2 UL36 proteins. The alignment is numbered according to the coordinates of the Md5 UL36 protein [42]. Vertical arrows are positioned over aligned amino acids with the greatest amount of positional diversity. Amino acids underlined with arrows indicate regions that contain reiterations. Solid arrows denote PA(P/S)KP(T/P/S), dotted arrows denote KPPP(D/A)PD(F/S)KP(S/T)PAPKS and dashed arrows denote (S/P)PA(S/P)KP. Solid black boxes 1, 2 and 3 illustrate regions in the alignment that differentiate CVI988 derivatives from more virulent strains. Abbreviations are the same as those in Table 3

```
                50        60        70        80        190       200       210       220       230       240
CVI988-INV   :  FTDHPTQKYKSAKAVSEDVSSTTRGGFTNKPRAKPG NENVKSKSERT------RRGGEIEIKSPDDPGSHRTHNPRTPATSRRHHSSARGYRGSDSE :243
CVI988-BP5   :  FTDHPTQKYKSAKAVSEDVSSTTRGGFTNKPRAKPG NENVKSKSERT------RRGGEIEIKSPDDPGSHRTHNPRTPATSRRHHSSARGYRGSDSE :243
CVI099-BAC   :  FTDHPTQKYKSAKAVSEDVSSTTRGGFTNKPRAKPG NENVKSKSERT------RRGGEIEIKSPDDPGSHRTHNPRTPATSRRHHSSARGYRGSDSE :243
RM-1         :  FTDHPTQKHKSAKAVSEDVLSTTRGGFTNKPRAKPG NENVKSKSERTTKSERTRRGGEIEIKSPDDPGSHRTHNPRTPATSRRHHSSARGYRSSDSE :249
R2/23        :  FTDHPTQKHKSAKAVSEDVSSTTRGGFTNKPRAKPG NENVKSKSERTTKSERTRRGGEIEIKSPDDPGSHRTHNPRTPATSRRHHSSARGYRSSDSE :249
CU-2         :  FTDHPTQKYKSAKAVSEDVSSTTRGGFTNKPRAKPG NENVKSKSERTTKSERTRRGGEIEIKSPDDPGSHRTHNPRTPATSRRHHSSARGYRSSDSE :249
JM/102W      :  FTDHPTQKHKSAKAVSEDVLSTTRGGFTNKPRAKPG NENVKSKSERTTKSERTRRGGEIEIKSPDDPGSHRTHNPRTPATSRRHHSSARGYRSSDSE :249
RB-1B-ADOL   :  FTDHPTQKHKSAKAVSEDVSSTTRGGFTNKPRAKPG NENVKSKSERTTKSERTRRGGEIEIKSPDDPGSHRTHNPRTPATSRRHHSSARGYRSSDSE :249
571          :  FTDHPTQKYKSAKAVSEDVSSTTRGGFTNKPRAKPG NENVKSKSERTTKSERTRRGGEIEIKSPDDPGSHRTHNPRTPATSRRHHSSARGYRSSDSE :249
GA           :  FTDHPTQKHKSAKAVSEDVSSTTRGGFTNKPRTKPG NENVKSKSERTTKSERTRRGGEIEIKSPDDPGSHRTHNPRTPATSRRHHSSARGYRSSDSE :249
549a         :  FTDHPTQKHKSAKAVSEDVSSTTRGGFTNKPRAKPG NENVKSKSERTTKSERTRRGGEIEIKSPDDPGSHRTHNPRTPATSRRHHSSARGYRSSDSE :249
595          :  FTDHPTQKHKSAKAVSEDVSSTTRGGFTNKPRAKPG NENVKSKSERTTKSERTRRGGEIEIKSPDDPGSHRTHNPRTPATSRRHHSSARGYRSSDSE :249
Md5          :  FTDHPTQKHKSAKAVSEDVSSTTRGGFTNKPRAKPG NENVKSKSERTTKSERTRRGGEIEIKSPDDPGSHRTHNPRTPATSRRHHSSARGYRSSDSE :249
Md11         :  FTDHPTQKHKSAKAVSEDVSSTTRGGFTNKPRAKPG NENVKSKSERTTKSERTRRGGEIEIKSPDDPGSHRTHNPRTPATSRRHHSSARGYRSSDSE :249
584a         :  FTDHPTQKHKSAKAVSEDVSSTTRGGFTNKPRAKPG NENVKSKSERTTKSERTRRGGEIEIKSPDDPGSHRTHNPRTPATSRRHHSSARGYRSSDSE :249
648a         :  FTDHPTQKHKSAKAVSEDVSSTTRGGFTNKPRAKPG NENVKSKSERTTKSERTRRGGEIEIKSPDDPGSHRTHNPRTPATSRRHHSSARGYRSSDSE :249
686          :  FTDHPTQKHKSAKAVSEDVSSTTRGGFTNKPRAKPG NENVKSKSERTTKSERTRRGGEIEIKSPDDPGSHRTHNPRTPATSRRHHSSARGYRSSDSE :249
```

Fig. 2. Multiple sequence alignments of variable regions within the UL49 proteins. Two regions (50–85 and 190–249) based on the coordinates of the Md5 UL49 proteins are presented [42]. Abbreviations are the same as those in Table 3

(Fig. 1). This region generally contains three types of repeat elements: PA(P/S)KP(T/S/P) – type A, (S/P)PA(S/P)KP – type B and KPPP(D/A)PD (F/S)KP(S/T) PAPKS – type C. The PAPKPP motif of type A repeats is only found in CVI988 derivatives and the mildly virulent CU-2 strain. These strains also contain the type B repeat motifs PPASKP and SPAPKP, which are not found in more virulent strains. Type C repeats of lesser virulent and attenuated strains contain serine residues in place of the more common threonine residues found in more virulent strains (Fig. 1 arrow at position 2797). Additional polymorphisms in three locations (boxes 1–3, Fig. 1) are largely responsible for the phylogenetic segregation of the CVI988 derivatives (Fig. 3a). These derivatives, along with CU-2, are of different lineage than all other strains including the attenuated strains RM-1 and R2/23.

As mentioned above, differences were also noted in the UL49 gene that encodes VP22. Alignments of these proteins (Fig. 2) reveal a sequence (TKSERT) missing in all CVI988 derivatives and a Gly^{245} polymorphism in contrast to the more common Ser^{245} of more highly virulent strains. A polymorphism at amino acid position 58 may also be of importance since Tyr^{58} occupies this position in all derivative of CVI988 and the mildly virulent strain CU-2. His^{58} is present in all virulent strains and the attenuated strains RM-1 and R2/23, which were derived from the virulent strains JM/102W and Md11, respectively. Like the partition seen with UL36 sequences, the UL49 dendrogram (Fig. 3b) shows grouping of the CVI988 derivatives and clustering of the vv and vv+ pathotypes.

Intergenic variation: 132-bp, a-like and US junction sequences

In addition to the polymorphisms in the UL36 and UL49 genes, two regions largely devoid of ORFs within the terminal repeat long and internal repeat long regions (TRL and IRL) were identified that vary among the 13 strains. These regions contain the so-called 132-bp tandem repeats [39] which have been reported to expand upon serial passage in cell culture concurrent with attenuation. We have determined the sequence and copy numbers of these

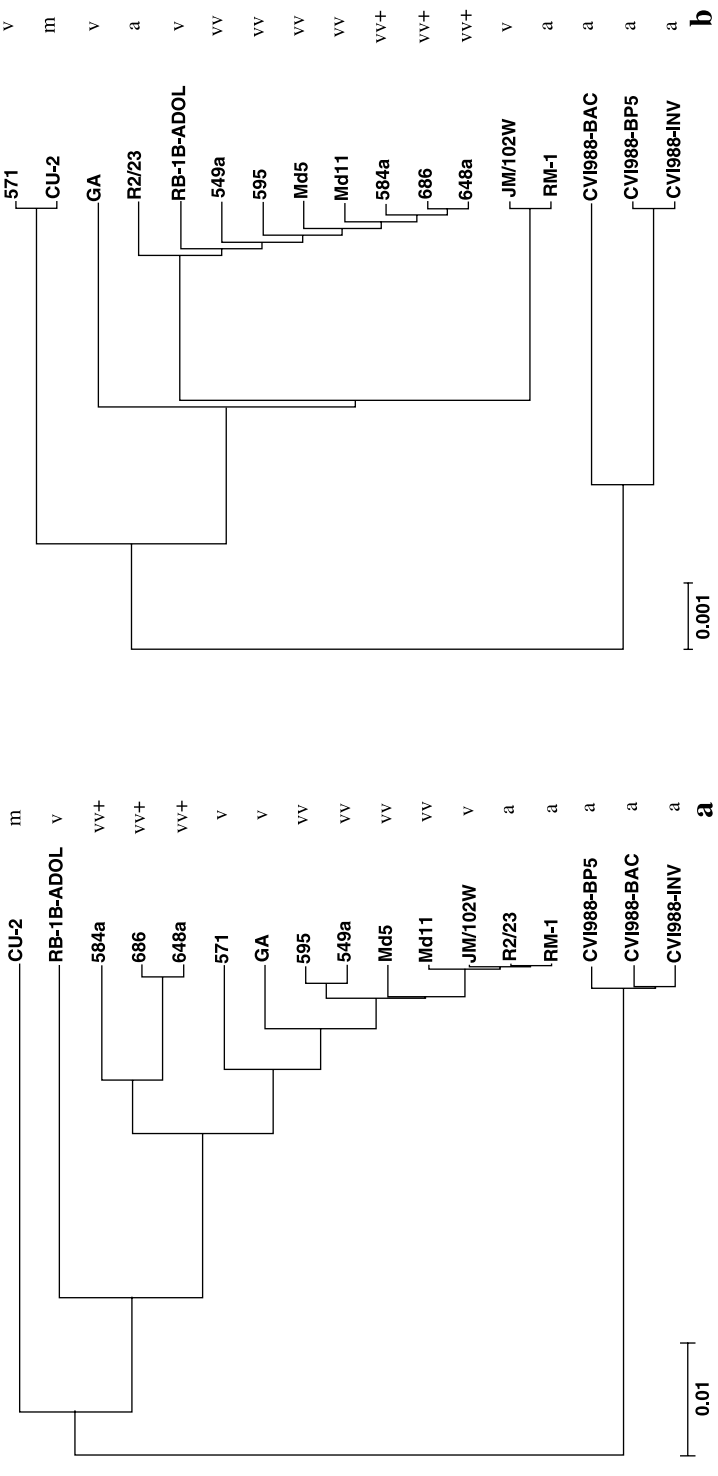

Fig. 3. Rooted dendrograms of the UL36 hypervariable regions (**a**) and UL49 proteins (**b**) from GaHV-2 strains. Pathotype classification of the GA, Md5, and Md11 strains of GaHV-2 is based on the reports by Lee et al., Niikura et al. and Tulman et al. [23, 28, 42]

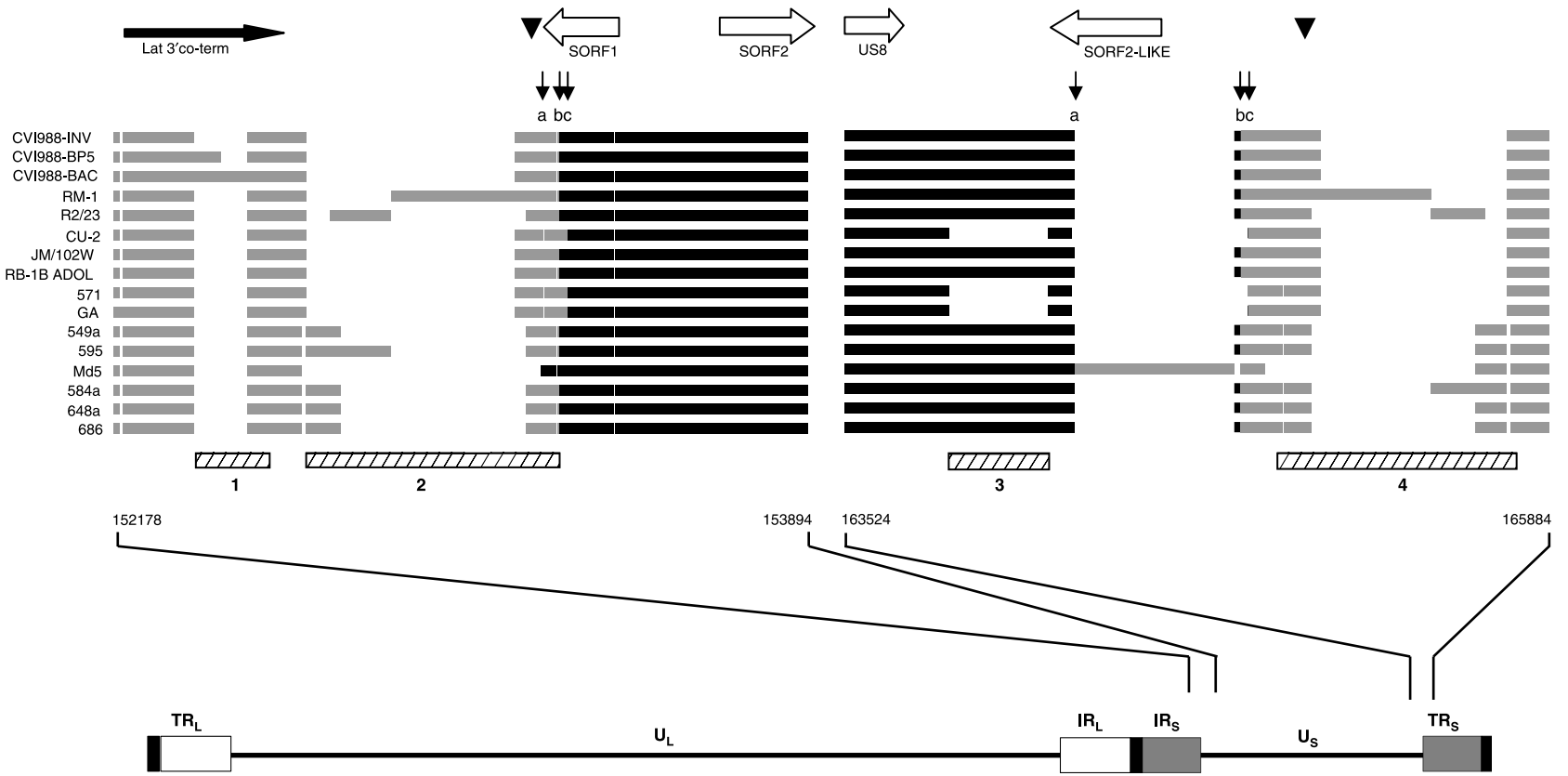

Fig. 4. Diagram of the IRS/US and US/TRS junctions of GaHV-2 strains. ORFs are denoted by open arrows. The black solid horizontal arrow indicates the position of co-terminal transcripts for latently infected lymphocytes [25]. Two large black vertical arrows indicate the position of the 65-bp sequence containing the consensus NF kappaB binding site. The position of the three paired junction boundaries (*a*, *b* and *c*) are indicated by arrows pointing down. Boxes 1, 2, and 4 denote regions of heterogeneity containing repeat sequences, and box 3 denotes deletions in strains CU-2, GA and 571

repeat regions for the 13 strains (Table 1) and found no correlation between copy numbers and pathotypes. However, we did find a SNP at position 98 that could collectively differentiate vv and vv+ (549a, 595, 584, 648a and 686) from the other three lesser pathotypes. Most of these latter strains contain reiterations of the 132-bp repeat containing cytosine at position 98. In contrast, all vv and vv+ strains examined in this study are heterozygous at position 98 for thymidine or cytosine.

Other regions of sequence heterogeneity flank the unique short region. As shown in Fig. 4, these junctions are inundated with repeat elements (boxes 1–4). Copies of a 115-bp repeat sequence (box 1) are present in some of the CVI988 derivatives: 3 copies in CVI988-BAC and 2 copies in CVI988-BP5. All other strains contain only one copy. Box 2 and its inverted complement box 4 also contain numerous repeats. Box 2 repeats are responsible for the heterogeneity in the COOH terminal sequences of the SORF1 proteins among the strains. Three translation products are predicted from the SORF1 genes with amino acid length groupings of 67, 88–89 and 104–105aa. Generally, vaccine strains contain the shortest versions. Downstream of SORF1 (3′ end of box 2 and 5′ end of box 4) are regions containing numerous transcription-factor-binding sites. Interestingly, there are differences in the number of binding sites among the strains. A 65-bp sequence containing an NF-kappaB binding site is absent in the genomes of vv and vv+ strains. This sequence is responsible for the branching of the vv/vv+ clade away from the less virulent strains (Fig. 5).

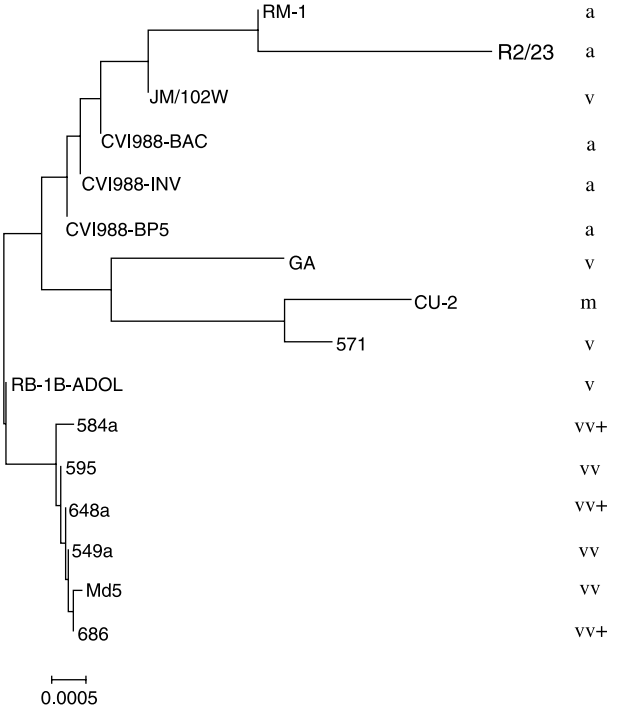

Fig. 5. Rooted dendrogram of box 4 nucleotide sequences from the US/TRS junction using the program Phylip

One surprising outcome of sequencing the IRS/US and US/TRS junctions was the lack of consensus regarding the placement of the exact junction; three junctions (a, b and c) were identified. One strain (Md5) contains it own unique junction (a) (Fig. 4). The junction between the IRS and US region for Md5 is similar to that of the others; however, unique short sequences encompassing the SORF2 ORF are also present in the TRS region. Due to this, the genome of Md5 contains two copies of SORF2 with varying carboxyl termini. Generally, most other strains have junction (b); however, strains CU-2, 571 and GA have a slightly different junction (c). Interestingly, these strains also contain a 485-bp deletion (box 3) in the unique short regions downstream of the US8 gene. Since this region in some strains has been reported to contain integrated retroviruses [21], it is not unreasonable to suggest that a retrotransposition mechanism may be responsible for the deletions.

Another set of regions that have remained fairly uncharacterized are *a*-like sequences found at the IRL/IRS and TRS/TRL junctions in the circular replicative intermediates. Because these sequences contain numerous repeat motifs including telomeric repeats, they are extremely difficult to clone and sequence. However, after much effort, we were able to determine the sequences of these regions from the 13 strains. Their sizes vary from 3453 to 5761 bp, with an average length of 4574 ± 719. Based on the multiple alignment of the nucleotide sequences (Fig. 6) we have identified sequences similar to the *pac-1* and *pac-2* sites of herpes simplex virus type 1 (HSV-1) (GGGGGGGGGGTGAAAATTT GGGGGG) and TTTTTTTTTATACAGTGTGT, respectively) which are bracketed by DR1 sites (GGCCGCGAGAGC). Internal to these sites are

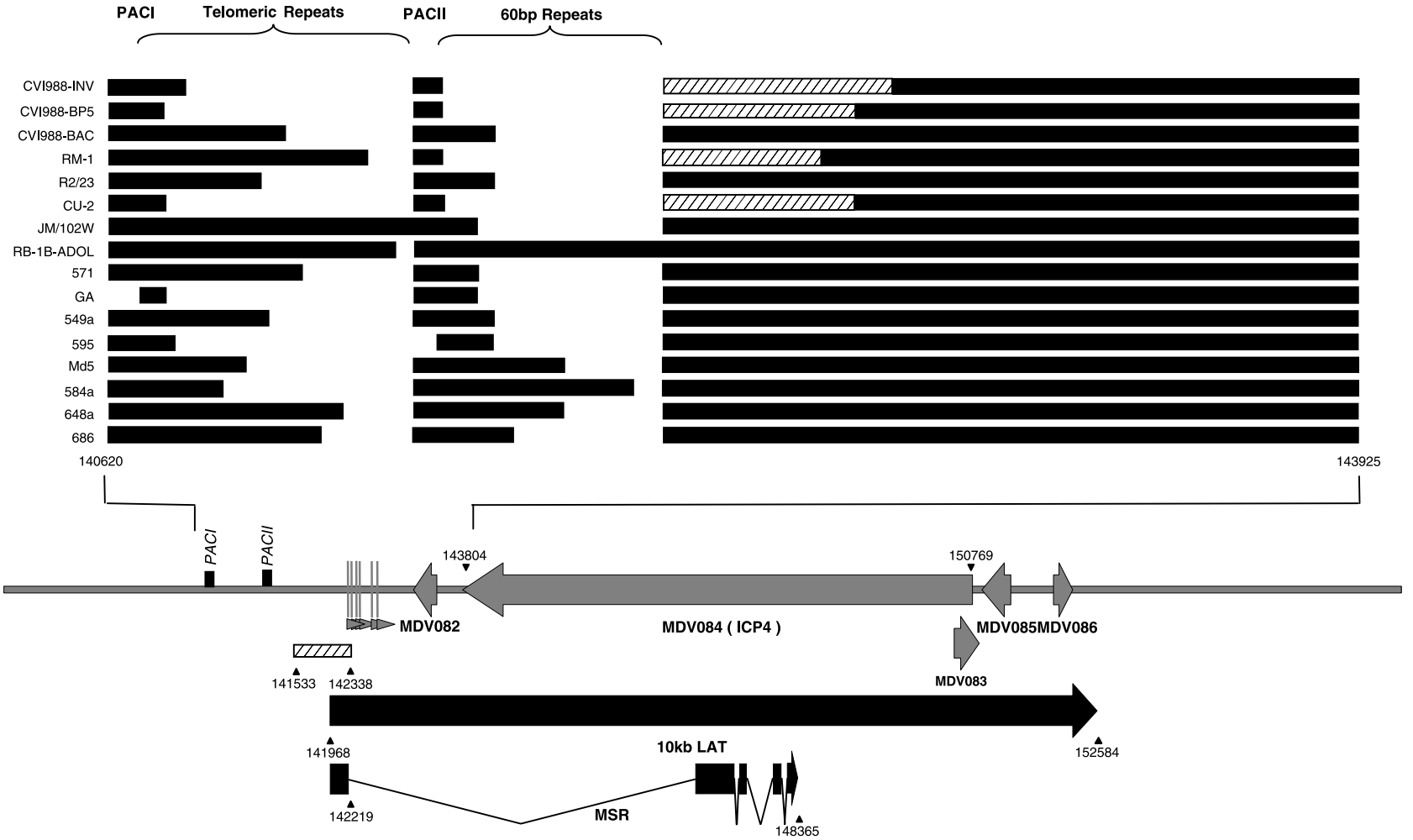

Fig. 6. Diagram of the *a*-like sequences of GaHV-2 strains and downstream sequences. Coordinates are based on the Md5 sequence [42]. Dashed boxes represent regions deleted in the attenuated strains CVI988-INV, CVI988-BP5, RM-1 and the mildly virulent strain CU-2. The extent of the deletions (nuc 141533 – nuc 142338) is illustrated in the enlarged portion containing the gene order

the telomeric repeats (GGGTTA) inundated with islands of the 13-bp repeats (GGGTTCAGGCCTA). Downstream of this region is a hypervariable region containing 60-bp repeats. In addition to variations in the telomeric and 60-bp repeat regions, four strains (CVI988-INV, CVI988-BP5, RM1 and CU2) contain deletions in a region immediately downstream of their 60-bp repeats. These deletions occur in the promoter region for the large LAT and spliced LAT [26] and in some strains encompass the first exon of the spliced LAT. In one strain (CVI988-INV) this deletion extends into the recently identified LAT microRNA region affecting MDV miR8 [3].

Discussion

The analysis of GaHV-2 genomes has been informative in pinpointing differences that may be associated with virulence. In this study, two genes (UL36 and UL49) and four regions (132-bp repeats, *a*-like sequences, IRS/US and US/TRS junctions) were characterized that differ among 13 strains that represent the 5 pathotypes. Phylogenetic analyses of these regions also suggest a relationship between the genotypes and pathotype classifications of the strains.

Sequence analysis of the UL36 genes indicated that CVI988 derivatives contain mutations and deletions in the proline-rich regions (boxes 1–3, Fig. 1). These polymorphisms appear to be unique to CVI988 variants and the mildly attenuated strain CU-2 but are not universally found in other attenuated strains (e.g. RM-1, R2/23). Intra-strain variations within box 2 (deletions in CVI988-BP5) were also observed. It is difficult to assess the contributable roles these deletions and SNPs may have in attenuation without a functional genomics approach. However, it is largely suspected that attenuation is multi-factorial and these polymorphisms – in at least the CVI988 derivatives and mildly virulent CU-2 strain – along with others might tip the scales in the direction of avirulence. Compounding these difficulties is the fact that little is known about the role of UL36 in virulence. The UL36 protein (~400 kDa) is the largest structural protein within the tegument of virions and interacts with other tegument proteins to mediate egress and re-envelopment [20, 43]. Interestingly, the polymorphisms within the UL36 proteins occur in proline-rich regions similar to SH3, WW and EHV1 domains that are involved in protein–protein interactions [18, 50]. Since an ubiquitin (Ub)-specific cysteine protease motif at the amino terminus has recently been identified in HSV-1 and GaHV-2 [51], it seems plausible that mutations in the proline-rich domains could affect retention of otherwise proteosome-targeted proteins and/or protein–protein interactions important for egress.

Another gene in the unique long region that differs among attenuated and oncogenic strains is UL49. This gene encodes the tegument phosphoprotein VP22, which has been shown to be essential for GaHV-2 propagation in cell culture [8, 9]. A short deletion of 18 bp (encoding TKSERT) was found in this gene within all variants of CVI988. This deletion corresponds to a domain that is rich in serine and threonine residues in the non-attenuated homologues. SNPs were also identified that partitioned to some degree along pathotypes. Relative to more highly virulent strains, CVI988 variants and the mildly virulent strain CU-2 contain a Tyr58His substitution and a Gly^{239} substitution instead of Ser^{245}. There is little sequence resemblance between the VP22 of GaHV-2 and its homologue in mammalian herpesviruses. However, like its HSV-1 counterpart, it is a phosphoprotein with DNA binding activity and displays similar functional properties [2, 8, 11, 33]. It remains to be determined whether residues in the TKSERT domain are phosphorylated. The contribution of UL49 polymorphisms to attenuation, like those of UL36, will require functional testing in which these genes are rescued into the genome of $UL49^{-}$ or $UL36^{-}$ knockout virulent mutants. Only then will the proof be provided that these two genes truly encode virulence factors. Small polymorphisms and even single amino acid substitutions can have a dramatic effect on the behavior of herpesviruses, as seen with mutations in the UL34.5 gene of HSV-1 resulting in a 100,000-fold reduction in neurovirulence [5, 32] and a single amino acid substitution in the polymerase gene of equine herpesvirus-1 allowing the differentiation of neurological and non-neurological strains [29].

Aside from the polymorphisms identified in the coding sequences of UL36 and UL49, other novel

insertions/deletions and SNPs were found in non-coding regions that are generally pathotype specific. These occur in the 132-bp repeats within the repeat long regions, the *a*-like sequence at the IRL/IRS junction and sequences flanking the US region. Over the last twenty years, the 132-bp repeat regions have received a considerable amount of attention due to their heterogeneity among MDV pathotypes. Recently, Silva and others have demonstrated that full loss of virulence requires additional changes in the genomes of recombinant Md5 viruses deleted in the 132-bp regions, indicates that, at best, these regions act as a marker for attenuation [38]. However, our sequencing research has indicated that the sequences of the 132-bp regions for the very virulent Md5 are different than all other vv and vv+ pathotypes examined. Within these genomes, a specific SNP (T98C) is present in at least one of the two tandem copies. Pathogenicity studies with vv and vv+ recombinant viruses containing deletions in these regions are needed to determine the significance of the SNP in virulence because of conflicting data in the literature suggesting that these repeats are essential for Ori-dependent replication in CEFs [17] and animal studies indicating similar replication rates between mutants deleted in the 132-bp repeats and wild-type parental viruses [36].

In addition to polymorphisms within the genes encoding tegument proteins and the 132-bp regions, sequence heterogeneity exists at both the IRS/US and US/TRS junctions in the 13 strains. These differences are likely the result of retrotransposition [16], as seen with the non-attenuated strains Md5 and JM [12, 21] and the attenuated strain RM-1 [15, 48]. Analyses of these junctions indicate that sequence heterogeneity exists in four regions (Fig. 4, boxes 1–4). Box 1 contains a 115-bp repeat sequence that is duplicated in some CVI988 derivatives. Box 3, which is in the unique short region, is missing in the strains CU-2, 571 and GA and may be important in defining the boundaries of the US region since these strains all have c-type junctions. Boxes 2 and 4 within the short inverted repeats are interesting for two reasons. First, these regions contain fragmented copies of repeat sequences that differ among the strains. Most strains contain mirrored images of these inverted repeat regions, but some (e.g. 595, Md5 and 584a) contain imperfect versions, indicating that sequence variations at the US junctions can be both inter- and intra-genomic. Secondly, a 65-bp sequence containing a consensus NF kappaB binding site is surprisingly missing in the genomes of the more virulent strains (595, 549a, 584a, 648 and 686). Jones et al. have reported that the attenuated strain RM-1, which was generated via retroviral insertion mutagenesis of the virulent JM/102W strain, contains an integrated provirus immediately upstream of the NF-kappaB binding site within the 65-bp sequence. Most significantly, a stable transcript initiating from the inserted LTR promoter has been reported to transcribe adjacent US genes. These regions may be important for transcriptional activation, and virulent strains lacking the 65-bp sequences may have deficiencies in gene expression. We are currently testing this hypothesis using microarray technologies.

Sequence analyses of the *a*-like sequences involved in the packaging of the viral genomes also generated unexpected results. Deletions of varying lengths in regions containing non-repeat (unique) sequences were identified in most vaccine strains. These deletions are downstream of the hyper-variable 60-bp repeat region within the promoter region of the LAT and in some strains extend to encompass a recently reported microRNA [3]. Little is known about the role of LAT or the functions of the viral encoded microRNAs.

Based on extensive comparative genomic studies, we have identified genes and noncoding regions within the genome of GaHV-2 that differ among the strains (attenuated vs non-attenuated). Phylogenetic analyses of the six regions reported here suggest that the attenuated strains derived from Rispens' original isolate all partition together, while other attenuated strains (RM-1 and R2/23) and to some degree the mildly virulent strain CU-2 partition largely with more highly virulent strains. We have identified regions (132-bp repeats) near the origins of replication that may be important for efficient replication and sequences in non-coding regions that may be important in the expression of surrounding genes. These comparative genomic studies need to be supplemented with a functional genomics approach involving the generation of viral recombinants and transcript profiling to more fully understand the significance of the reported findings.

Acknowledgements

We thank Jeremy Volkening for his excellent technical assistance and skills in bioinformatics and graphics. The authors would like to thank Drs. Laszlo Zsak and Cary Rue for input and critical evaluation of this report. This work was funded by the United States Department of Agriculture CRIS program (project number 6612-32000-043).

References

1. Afonso CL, Tulman ER, Lu Z, Zsak L, Rock DL, Kutish GF (2001) The genome of turkey herpesvirus. J Virol 75: 971–978
2. Blaho JA, Mitchell C, Roizman B (1994) An amino acid sequence shared by the herpes simplex virus 1 alpha regulatory proteins 0, 4, 22, and 27 predicts the nucleotidylylation of the UL21, UL31, UL47, and UL49 gene products. J Biol Chem 269: 17401–17410
3. Burnside J, Bernberg E, Anderson A, Lu C, Meyers BC, Green PJ, Jain N, Isaacs G, Morgan RW (2006) Marek's disease virus encodes MicroRNAs that map to meq and the latency-associated transcript. J Virol 80: 8778–8786
4. Calnek BW (2001) Pathogenesis of Marek's disease virus infection. In: Hirai K (ed) Current topics in microbiology and immunology, vol 255. Springer, Berlin Heidelberg New York Tokyo, pp 25–55
5. Chou J, Roizman B (1990) The herpes simplex virus 1 gene for ICP34.5, which maps in inverted repeats, is conserved in several limited-passage isolates but not in strain 17syn+. J Virol 64: 1014–1020
6. Corpet F (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 16: 10881–10890
7. Deng H, Dewhurst S (1998) Functional identification and analysis of cis-acting sequences which mediate genome cleavage and packaging in human herpesvirus 6. J Virol 72: 320–329
8. Dorange F, El Mehdaoui S, Pichon C, Coursaget P, Vautherot JF (2000) Marek's disease virus (MDV) homologues of herpes simplex virus type 1 UL49 (VP22) and UL48 (VP16) genes: high-level expression and characterization of MDV-1 VP22 and VP16. J Gen Virol 81: 2219–2230
9. Dorange F, Tischer BK, Vautherot JF, Osterrieder N (2002) Characterization of Marek's disease virus serotype 1 (MDV-1) deletion mutants that lack UL46 to UL49 genes: MDV-1 UL49, encoding VP22, is indispensable for virus growth. J Virol 76: 1959–1970
10. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797
11. Elliott G, O'Reilly D, O'Hare P (1999) Identification of phosphorylation sites within the herpes simplex virus tegument protein VP22. J Virol 73: 6203–6206
12. Endoh D, Ito M, Cho KO, Kon Y, Morimura T, Hayashi M, Kuwabara M (1998) Retroviral sequence located in border region of short unique region and short terminal repeat of Md5 strain of Marek's disease virus type 1. J Vet Med Sci 60: 227–235
13. Heinemeyer T, Wingender E, Reuter I, Hermjakob H, Kel AE, Kel OV, Ignatieva EV, Ananko EA, Podkolodnaya OA, Kolpakov FA, Podkolodny NL, Kolchanov NA (1998) Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic Acids Res 26: 362–367
14. Izumiya Y, Jang H, Ono M, Mikami T (2001) A Complete Genomic DNA Sequence of Marek's Disease Virus Type 2, Strain HPRS24. In: Hirai K (ed) Current topics in microbiology and immunology, vol 255. Springer, Berlin Heidelberg New York Tokyo, pp 191–221
15. Jones D, Brunovskis P, Witter R, Kung HJ (1996) Retroviral insertional activation in a herpesvirus: transcriptional activation of US genes by an integrated long terminal repeat in a Marek's disease virus clone. J Virol 70: 2460–2467
16. Jones D, Isfort R, Witter R, Kost R, Kung HJ (1993) Retroviral insertions into a herpesvirus are clustered at the junctions of the short repeat and short unique sequences. Proc Natl Acad Sci USA 90: 3855–3859
17. Katsumata A, Iwata A, Ueda S (1998) Cis-acting elements in the lytic origin of DNA replication of Marek's disease virus type 1. J Gen Virol 79 (Pt 12): 3015–3018
18. Kay BK, Williamson MP, Sudol M (2000) The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. FASEB J 14: 231–241
19. Kingham B, Zelnik V, Kopacek J, Majerciak V, Ney E, Schmidt C (2001) The genome of herpesvirus of turkeys: comparative analysis with Marek's disease viruses. J Gen Virol 82: 1123–1135
20. Klupp BG, Fuchs W, Granzow H, Nixdorf R, Mettenleiter TC (2002) Pseudorabies virus UL36 tegument protein physically interacts with the UL37 protein. J Virol 76: 3065–3071
21. Kost R, Jones D, Isfort R, Witter R, Kung HJ (1993) Retrovirus insertion into herpesvirus: characterization of a Marek's disease virus harboring a solo LTR. Virology 192: 161–169
22. Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 5: 150–163
23. Lee LF, Wu P, Sui D, Ren D, Kamil J, Kung HJ, Witter RL (2000) The complete unique long sequence and the overall genomic organization of the GA strain of Marek's disease virus. Proc Natl Acad Sci USA 97: 6091–6096
24. Lupiani B, Lee LF, Reddy SM (2001) Protein-coding content of the sequence of Marek's disease virus serotype 1. In: Hirai K (ed) Current topics in microbiology and immunology, vol 255. Springer, Berlin Heidelberg New York Tokyo, pp 159–190
25. McKie EA, Ubukata E, Hasegawa S, Zhang S, Nonoyama M, Tanaka A (1995) The transcripts from

the sequences flanking the short component of Marek's disease virus during latent infection form a unique family of 3′-coterminal RNAs. J Virol 69: 1310–1314
26. Morgan RW, Xie Q, Cantello JL, Miles AM, Bernberg EL, Kent J, Anderson A (2001) Marek's disease virus latency. In: Harai K (ed) Current topics in microbiology and immunology, vol 255. Springer, Berlin Hiedelberg New York Tokyo, pp 223–243
27. Nair V (2005) Evolution of Marek's disease – a paradigm for incessant race between the pathogen and the host. Vet J 170: 175–183
28. Niikura M, Dodgson J, Cheng H (2005) Direct evidence of host genome acquisition by the alphaherpesvirus Marek's disease virus. Arch Virol 151: 537–549
29. Nugent J, Birch-Machin I, Smith KC, Mumford JA, Swann Z, Newton JR, Bowden RJ, Allen GP, Davis-Poynter N (2006) Analysis of equid herpesvirus 1 strain variation reveals a point mutation of the DNA polymerase strongly associated with neuropathogenic versus nonneuropathogenic disease outbreaks. J Virol 80: 4047–4060
30. Osterrieder K, Vautherot J-F (2004) The genome content of Marek's disease-like viruses. In: Davison F, Nair V (eds) Marek's disease: an emerging problem. Elsevier, Amsterdam
31. Osterrieder N, Kamil JP, Schumacher D, Tischer BK, Trapp S (2006) Marek's disease virus: from miasma to model. Nat Rev Microbiol 4: 283–294
32. Perng GC, Mott KR, Osorio N, Yukht A, Salina S, Nguyen QH, Nesburn AB, Wechsler SL (2002) Herpes simplex virus type 1 mutants containing the KOS strain ICP34.5 gene in place of the McKrae ICP34.5 gene have McKrae-like spontaneous reactivation but non-McKrae-like virulence. J Gen Virol 83: 2933–2942
33. Phelan A, Elliott G, O'Hare P (1998) Intercellular delivery of functional p53 by the herpesvirus protein VP22. Nat Biotechnol 16: 440–443
34. Retief JD (2000) Phylogenetic analysis using PHYLIP. Methods Mol Biol 132: 243–258
35. Rispens BH, van Vloten H, Mastenbroek N, Maas HJ, Schat KA (1972) Control of Marek's disease in the Netherlands. I. Isolation of an avirulent Marek's disease virus (strain CVI 988) and its use in laboratory vaccination trials. Avian Dis 16: 108–125
36. Silva RF, Gimeno I (2007) Oncogenic Marek's disease viruses lacking the 132 base pair repeats can still be attenuated by serial in vitro cell culture passages. Virus Genes 34: 87–90
37. Silva RF, Lee LF, Kutish GF (2001) The genomic structure of Marek's disease virus. In: Hirai K (ed) Current topics in microbiology and immunology, vol 255. Springer, Berlin Hiedelberg New York Tokyo, pp 143–158
38. Silva RF, Reddy SM, Lupiani B (2004) Expansion of a unique region in the Marek's disease virus genome occurs concomitantly with attenuation but is not sufficient to cause attenuation. J Virol 78: 733–740
39. Silva RF, Witter RL (1985) Genomic expansion of Marek's disease virus DNA is associated with serial in vitro passage. J Virol 54: 690–696
40. Spatz SJ, Petherbridge L, Zhao Y, Nair V (2007) Comparative full-length sequence analysis of oncogenic and vaccine (Rispens) strains of Marek's disease virus. J Gen Virol 88: 1080–1096
41. Spatz SJ, Silva RF (2006) Polymorphisms in the repeat long regions of oncogenic and attenuated pathotypes of Marek's disease virus 1. Virus Genes 35: 41–53
42. Tulman ER, Afonso CL, Lu Z, Zsak L, Rock DL, Kutish GF (2000) The genome of a very virulent Marek's disease virus. J Virol 74: 7980–7988
43. Vittone V, Diefenbach E, Triffett D, Douglas MW, Cunningham AL, Diefenbach RJ (2005) Determination of interactions between tegument proteins of herpes simplex virus type 1. J Virol 79: 9566–9571
44. Witter RL (1997) Increased virulence of Marek's disease virus field isolates. Avian Dis 41: 149–163
45. Witter RL (2001) Marek's disease virus vaccines-past, present and future (chicken vs. virus-a battle of the centuries). In: Schat KA, Morgan RM, Parcells MS, Spencer JL (ed) Current progresson Marek's disease research. Proc, 6th International Symposium on Marek's disease. American Association of Avian Pathologists, Kennett Square, PA
46. Witter RL (2001) Protective efficacy of Marek's disease vaccines. In: Hirai K (ed) Current topics in microbiology and immunology, vol 255. Springer, Berlin Hiedelberg New York Tokyo, pp 57–90
47. Witter RL, Calnek BW, Buscaglia C, Gimeno IM, Schat KA (2005) Classification of Marek's disease viruses according to pathotype: philosophy and methodology. Avian Pathol 34: 75–90
48. Witter RL, Li D, Jones D, Lee LF, Kung HJ (1997) Retroviral insertional mutagenesis of a herpesvirus: a Marek's disease virus mutant attenuated for oncogenicity but not for immunosuppression or in vivo replication. Avian Dis 41: 407–421
49. Witter RL, Nazerian K, Purchase HG, Burgoyne GH (1970) Isolation from turkeys of a cell-associated herpesvirus antigenically related to Marek's disease virus. Biken J 13: 53–57
50. Zarrinpar A, Bhattacharyya RP, Lim WA (2003) The structure and function of proline recognition domains. Sci STKE 2003: RE8
51. Zhao Y, Smith L, Petherbridge L, Baigent S, Nair V (2006) Investigation on the deubiquitinating enzyme activity of Marek's Disease Virus UL36 homolog. In: Parcells MS, Morgan RW, Burnside J, Schmidt CJ, Bernberg EL, Eppler E (eds) 4th International Workshop on the Molecular Pathogenesis of Marek's Disease Virus, pp 27